

CHAPTER 12

PORTFOLIO ASSESSMENT: SCORING AND STUDENT PERFORMANCE

THE PLACE OF PORTFOLIO ASSESSMENT IN KIRIS

Portfolios occupied a key place in KIRIS, both as a means of assessment that directly tapped student work in classrooms, and as a means for supporting educational improvement in classrooms, schools, and districts. Since the contents of the portfolios arose from students' classroom work, the portfolio was the assessment component that most clearly reflected local curriculum and instruction. In concept, students developed portfolios over long periods, months and perhaps years. Because students had the opportunity to revise their portfolio entries, with feedback from teachers and peers, the assessment portfolio may reasonably be viewed as the student's "best work."

In many respects portfolios were the portion of the KIRIS assessment that most directly and comprehensively supported educational reform because of the strong connection to students' classroom experiences and strong involvement of teachers. For that reason, the KIRIS portfolio activities included extensive professional development opportunities, which local schools employed as a powerful means to support teachers' professional development and school improvement. Both Writing and Mathematics, (Mathematics Portfolio was deleted mid-way through the third accountability cycle), employed a trainer-of-trainers model to deliver scoring training throughout the state. In addition, Regional Consultants provided professional development and informal teacher support throughout the year.

Writing Portfolios were fully implemented in the first KIRIS Accountability Cycle, and continued in Accountability Cycles 2 and 3; Mathematics Portfolios and Alternate Portfolios were introduced during Accountability Cycle 1, with Mathematics Portfolios removed from accountability after 1996. All KIRIS portfolios were scored locally to allow each school to observe all of the material included in the portfolios, material that often exceeded the scoring criteria. Reliance on local scoring required training and practice as well as alignment between portfolio requirements and local instruction. The portfolio development and scoring process also assumed considerable content knowledge on the part of teachers. Although external scoring could provide summary data in the form of scores and standardized notations (see description of scoring analysis), only local portfolio scoring allowed discussion of the best ways to modify instruction based on assessment data that directly reflected classroom practices. Extensive professional development was provided throughout the state to support scoring accuracy and the alignment of instruction with the portfolio assessment criteria.

All KIRIS portfolio assessments were standardized in the following ways: (1) all required the use of a scoring guide accompanied by several benchmarks or exemplars for each

score point, (2) portfolio content requirements prescribed the number and types of required entries, portfolios that were incomplete were scored as Novice (zero points), and (3) the state provided every school district with complete scoring training materials for each accountability grade, including detailed rules for portfolio preparation, to assure that the student completed the work in each portfolio.

MONITORING THE SYSTEM

Portfolio development and scoring were monitored in several ways. Early in the year, a Code of Ethics was distributed to all schools describing the limits on a teacher's comments or modifications of a student's portfolio entries. When the school assigned portfolio scores were submitted to KDE, the principal was required to submit a signed assurance statement confirming that appropriate portfolio development practices were observed. Accusations or complaints of inappropriate practices were investigated by KDE and penalties applied, if warranted. In addition, each student portfolio included a statement signed by the student attesting that s/he completed all portfolio entries. If local scoring personnel discovered plagiarism or an audit revealed it, the entry was removed, making the portfolio incomplete. Incomplete portfolios received a score of zero.

The public accountability and high stakes associated with the assessment system necessitated continued monitoring of portfolio development and scoring. Two different types of activities were used to provide schools with feedback on their scoring effectiveness, the scoring audit and the scoring analysis. An audit was a formal study of local scoring accuracy. The audit had several purposes: 1) to monitor accuracy of scoring throughout the system in order to plan statewide training and allocation of resources; 2) to correct inaccurate scores assigned locally; and 3) to verify exceptional score gains. KDE defined the sample of schools, usually in a manner that allowed the results to be generalized beyond the group of participating schools. All schools selected for an audit had to participate; they were required to submit all portfolios for re-scoring; and locally assigned portfolio scores and the resulting cognitive index were changed as a result of the audit.

The other type of activity was scoring analysis. Scoring analysis sessions were designed to provide schools with information about their scoring accuracy without the risk of changes to their school accountability scores. The analysis used many of the audit scoring procedures, but neither scores nor the accountability index were changed. The emphasis was upon providing commentary, based on examination of a school set of portfolios, which helped schools plan future instruction and professional development. School participation was voluntary.

After five years of portfolio assessment, two main issues continued to surround the use of portfolios in Kentucky's assessment system; 1) the level of scoring accuracy achieved by Kentucky teachers, and 2) the impact of portfolios on instructional practice. The following sections present the rationale for the design of each portfolio, information about the scoring reliability, and the instructional impact of design and reliability during

Accountability Cycle 3 (1995 through 1998),¹ and a discussion of related issues.

THE KIRIS WRITING PORTFOLIO

RATIONALE AND DESIGN The Kentucky Writing Portfolio assessed student writing directly (at grades 4, 8, and 12 for the first two years of Accountability Cycle 3, and grades 4, 7, and 12 for the last two years) by examining a collection of a student's written products. The structure of the Writing Portfolio and the holistic scoring guide encouraged teachers to provide instruction focused on teaching students to communicate effectively and to provide grammar, punctuation, and spelling instruction through these authentic writing experiences. Because of the holistic scoring of Writing Portfolios, there was some concern that there could be a possible loss of student spelling and grammar skills. A research study was conducted to see if these concerns were valid. Archived writing test responses for grade 4 from 1993 and 1996 were scored for errors in spelling, capitalization, punctuation, and subject/verb agreement. Research results indicated improvement on the holistic writing scores and spelling from 1993 to 1996. A composite index composed of capitalization, punctuation, and subject/verb agreement skills remained stable. However, boys and girls scores differed. Boys showed improvement in spelling and on the capitalization, punctuation, and subject/verb agreement index. Girls did not improve on these indexes. A more complete report on these findings can be seen in Appendix O.

A committee of Kentucky English/Language Arts educators originally designed the portfolio. This committee discussed the traditional writing experience of Kentucky students and found that most instruction focused on isolated grammar and very confined writing experiences (i.e., reports, essays, research papers). Using the writing Academic Expectation as their guide (that all students should write for multiple purposes in multiple forms for a variety of audiences), the committee structured the contents of the portfolio to include broad categories of writing that consciously excluded reports, academic essays, and research papers. Instead, the committee created a structure that required other types of writing:

- Personal experience writing;
- Imaginative writing;
- Reflective writing; and,
- Trans-active writing for real-world purposes and audiences.

In addition to this purposeful design of the portfolio contents, the criteria for assessment were selected and scoring tools designed with these instructional focus changes clearly in mind. While the committee believed that mastery and assessment of mechanics remained critical, they also identified several more critical criteria that traditionally had been less evident in writing instruction and assessment in Kentucky (e.g., organizational skills, idea development, and focus on purposes and audiences). Finally, the committee

¹ Information about activities occurring prior to 1993 may be found in the Cycle 2 *Technical Manual*.

selected six main criteria for assessing the quality of student writing. These criteria were applied holistically to produce a single final judgment for a complete portfolio. The criteria follow:

- Purpose/Audience Awareness;
- Idea Development/Support;
- Organization;
- Sentence Structure and Variety;
- Language (Word Choice and Usage); and,
- Correctness (Spelling, Punctuation, and Capitalization).

The committee believed that these portfolio content requirements and assessment criteria would provide teachers with guidelines for more balanced writing instruction, consistent with the national movement toward more process-centered instruction.

PROFESSIONAL DEVELOPMENT

The Kentucky Writing Program (KWP) supported a wide variety of professional development experiences including portfolio scoring training as well as workshops and consultation focused primarily on classroom strategies for developing student writing skills. Since the introduction of the Writing Portfolio Assessment, the KWP had supported classroom teachers through a tiered training system that relied on a design committee to train local trainers; who then delivered portfolio development and scoring strategies to the other teachers in their school. Each year, these local trainers received two rounds of professional development: one focused on the generation of portfolio entries and one focused on scoring to state standards. Print materials, video training provided through statewide educational television, and local level workshops provided by the KWP Regional Consultants augmented these sessions. In addition, Regional Writing Consultants worked with local districts and schools upon request to provide individually tailored professional development experiences focused on a variety of Writing Portfolio related topics such as portfolio analysis, technical writing, personal experience writing, reflective writing, writing across the curriculum, development of writing workshop classrooms, and designing appropriate assignments focused on real-world purposes and audiences.

1996 WRITING PORTFOLIO SCORING AUDIT²

The 1996 Writing Portfolio scoring audit was the second to be carried out, not counting the small voluntary audit conducted in 1994. The results and the consequences for schools were very different from the previous audit. The first Writing Portfolio scoring audit was held in 1993 (for details see the *KIRIS Accountability Cycle 1 Technical Manual*). By legislative directive, the 1993-writing audit allowed schools the choice of keeping their original scores or accepting revised scores based on the generally lower

² For complete procedures and full results of the audit, see *1996 Writing Portfolio Audit: Final Report*

audit results. Most schools chose to use the scores that they had assigned to the portfolios to compute both their writing cognitive index and the KIRIS accountability index.

In 1996, a second audit monitored scoring patterns for 100 schools and adjusted scores for those schools found to be scoring portfolios inaccurately. The audit results were reported to individual schools; and the audit scores were used to adjust the writing cognitive index where necessary. Those conducting the audit observed that local scoring was much more accurate than in 1993.

The scoring contractor conducted the 1996 audit with a small proportion of the audit portfolios re-scored by experienced Kentucky teachers to assure that the contractor had applied the scoring standards correctly. Audit scorers, trained in the use of the Kentucky Scoring Guides, qualified by scoring a set of portfolios to a predetermined level of accuracy (six of nine correct and no more than three points difference on all nine) and were monitored continuously for scoring accuracy. Eight of nine scorers qualified at the elementary level, seven of seven at the middle level, and eight of thirteen at the high school level. Since there was no provision for retraining, the non-qualifiers were dismissed. Similarly, any audit scorer who failed to maintain the required level of scoring accuracy was discharged.

One hundred schools were selected for auditing. (Because two of the selected schools had been closed, the final group of schools audited numbered 98.) Half, designated as the random schools, were selected at random, providing a sample of schools from which to infer statewide scoring accuracy rates. The other half, referred to as the purposeful schools, were chosen using a formula that identified schools with Writing Portfolio scores that were very high or very low relative to test scores in other content areas. Further information about the conducting of the audit is available in 1996 Writing Portfolio Audit: Final Report, by Amy Aubrey, available from the Kentucky Department of Education, Office of Assessment and Accountability.

THE 1997 AND 1998 WRITING PORTFOLIO AUDITS

After the 1996 administration of the Writing and Mathematics Portfolios, there were major changes in the system. The Mathematics Portfolio was placed in a two-year research and development status, and then was not returned to the accountability system at the end of that status in 1998. This was discussed more completely in Chapter 7, and will not be further developed here.

Following the 1996 testing, the Writing Portfolio was moved from grade 8 to grade seven. Consequently, the Writing Portfolio was removed from the accountability for a two-year (1997 and 1998) period of adjustment to the new grade level. Therefore, Writing Portfolio audits were conducted only at grades 4 and 12 in 1997 and 1998.

The three expressed purposes of the 1996 audit (see page 12-2) were more explicitly stated as five objectives for the 1997 and 1998 audits. The objectives were to 1) adjust

Chapter 12

Portfolio Assessment: Scoring And Student Performance

discrepant scores; 2) encourage accurate scoring; 3) provide data for training; 4) provide data about statewide accuracy, and 5) establish the audit as an annual expectation. The selection process remained two tiered, with a purposeful selection of 50 schools and a random selection of 50.

The professional scoring teams of the contractors in Dover, New Hampshire, and Minnetonka, Minnesota, scored the 1997 and 1998 audited portfolios. Most of the scorers were college graduates; with many of them retired teachers or businesspersons. In 1997, scorers were trained on nine pre-scored portfolios. Those who differed by more than three performance levels total, or had less than six correct were not allowed to score. During the scoring, table leaders double scored a portfolio at least four times per day to verify consistency and accuracy. As additional quality control, pre-scored portfolios were seeded into the portfolios to be scored on a daily basis. After the portfolios were scored, the audit score was compared to the school assigned score. If the scores agreed, the school assigned score was final. Those where the auditor disagreed were read by yet another auditor. Whichever two scores agreed determined the score for the portfolio, meaning that at times the school score stood, and for others the auditors' scores stood.

In 1998, KDE and the contractor both participated in training the scorers. A packet of five portfolios was used for training. Both KDE and the contractor monitored the scoring throughout the period of days required to score all the portfolios. All readers scored three sets of five portfolios to qualify. The standard was 80percent accuracy in at least two of the packets. To maintain consistency team leaders read 20percent of the portfolios in each pack read by scorers. Both audit scores were recorded for checking against the school assigned score. Scoring directors and KDE personnel read 20percent of the portfolios scored by the team leaders. In addition, readers scored two pre-determined quality control portfolios that were seeded in the packets, one in the morning and one in the afternoon. Readers were required to maintain a 75percent agreement rate with team leaders. Failure to reach this rate for three consecutive days caused the dismissal of six readers during scoring.

The results of the audit are summarized in table 12-1. Two measures of scoring accuracy are presented, the exact agreement between the portfolio scores assigned by the school and those assigned by the audit, and the magnitude of the difference in the Writing Cognitive index (WCI), the mean scores of all portfolios on a scale of 0 to 140 points. In 1996, the purposeful group exhibited slightly lower exact agreement with the audit scores with 73 percent of their portfolio scores confirmed by one or more audit readers. For the random group, 77 percent of the locally assigned scores were confirmed. In 1997, that pattern reversed with 77 percent of the purposeful agreeing, compared to 74 percent of the random. In 1998, the pattern reversed again with 69 percent of the purposeful group agreeing and 75 percent of the random group agreeing. Across schools, the 1996 mean change in the WCI was nearly twice as large for the purposeful group as for the random group. That pattern was reversed in 1997 and then reversed again in 1998. The audit adjusted the WCI for the purposeful and random groups downward all three years with the amount of adjustment increasing over the

three-year span. The WCI was determined by assigning a score of 0 for all portfolios scored blank, incomplete, or novice. Values of 40, 100, and 140 points are assigned to portfolios rated as apprentice, proficient or distinguished, respectively. The WCI was then computed as the arithmetic mean of all portfolio scores.

TABLE 12-1
SUMMARY OF CYCLE 3 AUDIT RESULTS

	1995	1996		1997		1998	
Sample	No Audit	Original Audit Score Agreement	Mean Change in Writing Cognitive Index	Original Audit Score Agreement	Mean Change in Writing Cognitive Index	Original Audit Score Agreement	Mean Change in Writing Cognitive Index
Purposeful	N/A	73%	-10.87	77%	-7.00	69%	-14.12
Random	N/A	77%	-5.09	74%	-9.49	75%	-9.68

It is important to note that scoring accuracy was not uniform across schools. Nearly 70 percent of the schools in the 1996 audit assigned scores in exact agreement with the audit for 70 percent or more of their portfolios. In 1997, this increased to 80 percent of the schools scoring with 70 percent or more agreement. While there was a drop in 1998, the percentage of schools scoring at least 70 percent accurately remained at 70 percent. This is a remarkable change from 1993 when no schools scored portfolios at this level of accuracy. Table 12-2 provides comparison data based on the exact agreement between locally assigned scores and audit scores.

The first audit of Writing Portfolios, conducted in 1993, demonstrated that schools were significantly over-estimating student performance. On average, the audit-score WCI was nearly 40 points lower than the school-score WCI. The 1996 audit presents very different results.³

³ It is important to note that the selection process used for this first auditing activity was solely purposeful (e.g., all schools were selected for auditing due to apparent discrepancies in score gains in writing relative to other content areas). Therefore, comparisons between this early audit and the recent audit can only be made within the pool of purposefully selected schools. (For more information about the 1993 audit see the *Cycle 1 Technical Manual*.)

TABLE 12-2 1995 THROUGH 1998 AUDITS: LEVELS OF EXACT AGREEMENT				
	1995 ¹	1996	1997	1998
Exact Agreement	Total Number of Schools	Total Number of Schools	Total Number of Schools	Total Number of Schools
90% or more		10	16	8
80 - 89%		33	30	31
70 - 79%		25	34	31
60 - 69%		18	9	13
50 - 59%		8	4	10
Less than 49%		4	7	7

¹No writing portfolio audit was conducted in 1995.

TABLE 12-3 COMPARISON OF CYCLE 3 WRITING PORTFOLIO CHANGES TO WRITING COGNITIVE INDEX						
Audit Year	Type of Sample	Writing Cognitive Index	Grade 4	Grade 8	Grade 12	Total Schools
1995 ¹	Purposeful	Original				
		Audited				
	Random	Original				
		Audited				
1996	Purposeful	Original	54.6	26.3	55.4	44.0
		Audited	46.0	12.9	42.0	33.1
	Random	Original	40.4	25.8	29.4	31.9
		Audited	35.2	21.5	23.9	26.8
1997 ²	Purposeful	Original	45.9		46.2	46.0
		Audited	40.3		33.4	39.0
	Random	Original	47.4		52.5	49.3
		Audited	42.5		35.5	39.8
1998 ²	Purposeful	Original	54.3		54.7	54.4
		Audited	41.6		33.9	40.2
	Random	Original	43.2		47.9	45.5
		Audited	35.7		36.2	35.9

¹No writing portfolio audit was conducted in 1995.

²No audit conducted during the two-year transition period when the writing portfolio moved to grade seven.

The results for 1996 demonstrated substantial improvement in local scoring accuracy over 1993. While improvement stabilized from 1996 through 1998, the difference between mean audited scores and the original mean scores demonstrated the need for continued improvement. Table 12-3 demonstrates even more clearly that audits lowered portfolio scores, and that purposefully selected schools tend to be lowered more than randomly selected schools.

STUDENT PERFORMANCE

The data indicates that caution about conclusions is necessary. Table 12-4 shows the state mean score by grade level for the Writing Portfolio over the four years of Cycle 3. There was a little change downward in Writing Portfolio scores at the middle level and only modest gains in elementary and high school scores. The consistent over-scoring at the local level probably masks real gains in student performance.

TABLE 12-4				
STATEWIDE WRITING PORTFOLIO MEAN SCORES				
GRADE	1995	1996	1997	1998
4	40.3	39.3	43.8	43.8
8	31.8	27.5	28.1	28.8
12	38.6	38.9	42.1	43.4

USING WRITING PORTFOLIO SCORING ANALYSIS TO PROVIDE INSTRUCTIONAL FEEDBACK

During the summers of 1994 and 1995, all Kentucky schools had the opportunity to submit portfolios for scoring analysis. The purpose of portfolio scoring analysis was twofold: to provide a school with informal feedback on their scoring accuracy and to provide recommendations about instructional practices and future professional development based on a review of portfolio contents. Also, the scoring analysis session modeled an activity that schools could carry out independently. School participation was completely voluntary. Although the results from this activity cannot be generalized to draw firm conclusions about typical instructional practices statewide, the scoring analysis does provide information about widespread teacher understanding of best practices in writing instruction.

The scorers for the scoring analysis were Kentucky teachers trained and qualified to score and analyze school sets of portfolios. Scorer accuracy was monitored throughout the session. A panel of six readers evaluated school sets of 20-30 portfolios, scoring student work, and elucidating instructional strengths on the basis of the writing products in those portfolios. The panel conferred to draft recommendations concerning future instructional foci and professional development for the school.

RECOMMENDATIONS. The 1995 scoring analysis, the portion of the analyses

included in Accountability Cycle 3, examined more than 10,000 portfolios. A team of six readers worked together to review and discuss instructional evidence present in the portfolios from a particular school. Feedback was provided by noting extraordinary performance on ten descriptors that are based on the six assessment criteria of the scoring guide. After discussing the school set of portfolios, the group reached consensus on which of ten possible descriptors should be checked as strength or a need. (A neutral result was not noted.) A single report was prepared for each school. Table 12-5 summarizes the results across all participating schools.

TABLE 12-5 INSTRUCTIONAL ANALYSIS STRENGTHS AND NEEDS (1995 SCORING ANALYSIS)								
Instructional Analysis Notations	Grade 4 (n=20)		Grade 8 (n=10)		Grade 12 (n=4)		Total (34 schools)	
	Strength	Need	Strength	Need	Strength	Need	Strength	Need
Establishing Focused Purposes	20%	64%	14%	68%	23%	63%	19%	65%
Writing for Authentic Audiences	15%	46%	11%	54%	14%	58%	14%	50%
Employing Suitable Voice/Tone	59%	3%	40%	11%	53%	9%	53%	6%
Developing Ideas	14%	58%	3%	60%	10%	59%	11%	59%
Including Supporting Details	5%	54%	1%	65%	10%	23%	5%	51%
Organizing Ideas Logically	33%	25%	32%	24%	47%	12%	35%	23%
Using Effective Transitions	4%	26%	1%	22%	2%	6%	3%	21%
Constructing Sentences	24%	9%	16%	24%	14%	5%	20%	12%
Using Language Effectively	16%	8%	7%	27%	17%	15%	14%	14%
Editing for Correctness	35%	18%	46%	9%	23%	32%	35%	18%

Areas that were noted as needs at all grade levels for more than 50 percent of the participating schools were establishing focused purposes, and developing ideas. In addition, writing for authentic audiences and including supporting details were needs for more than 50 percent of the participating schools at two of the three grade levels. Strength was most frequently noted in the area of employing suitable voice/tone. The scoring analysis readers also recommended the elements of the writing process upon which they thought the school should focus to improve future portfolio results, as well

as, the types of portfolio entries on which schools should focus resources.

SUMMARY

Writing Portfolio activities during Accountability Cycle 3 can be summarized in the form of three questions.

Is scoring accuracy improving? From 1993 to 1996, there was dramatic improvement, and from 1996 to 1997, there was a modest improvement. In 1998, the scoring accuracy seemed to revert to the 1996 level, although over 70 percent of the schools agreed on over 70 percent of the portfolios. The 1996 audit documented substantial improvement in local scoring accuracy from 1993. In 1993 only 23 percent of the audit-assigned scores agreed with school-assigned scores, and the audit average index was 40 points lower than the school average index for purposefully audited schools. Three years later, after significant training and professional development efforts, agreement between school-assigned scores and audit-assigned scores rose to 74 percent, and the average index difference decreased to 10 points for purposefully audited schools. For schools selected at random (providing a picture of overall statewide accuracy), the audit confirmed 76percent of the school assigned scores with an average index difference of only five points on a 140-point index. When 1998 is compared to 1996, as demonstrated in table 12-4, statewide portfolio scores rose approximately 3.5 points at grade 4 and grade 12, and a smaller amount at the middle level. This suggests that the quality of portfolios audited improved gradually. Table 12-2 reveals, however, that the number of audited schools able to score 70percent or more of their portfolios in agreement with the auditors moved only from 68percent to 70percent, while the schools that scored below 60percent in agreement moved from 12percent to 17percent. Thus, the continuing efforts to train teachers to score accurately are necessary so that poor scoring does not call into question the gains.

Have writing portfolio results improved? When scoring inaccuracy is the product of inflated performance judgments, improved accuracy creates the appearance of reduced scores. Therefore, the stable KIRIS results of the last three years of Accountability Cycle 3 may be interpreted as an increase in actual performance. In fact, audit data for purposefully selected schools (the sole sample available over time) demonstrated a slight decline in scoring accuracy, but small increases in student performance at grades 4 and 12. The results are mixed enough that this question remains unanswered. One additional indicator, that at least the top schools are actually improving their scores, is that from 1996 to 1998 the purposefully selected (highest scoring) average index score at the grade 4 level rose from 46.0 to 54.3. At grade 12 the average index score rose from 42.0 to 54.6. These increases occurred while scoring accuracy remained relatively constant and thus represent genuine improvement.

Can portfolios be accurately analyzed for instructional strengths and needs?

Independent readings of school samples of portfolios produced similar recommendations for future instructional emphasis and professional development. The consistency of the conclusions drawn by independent readers suggested that this

analysis procedure could be very helpful to schools that want to plan effective instructional improvement based on Writing Portfolio contents.

SELECTED REPORTS ON WRITING PORTFOLIOS AVAILABLE FROM KDE

1994-95 Writing Portfolio Scoring Analysis Report

1995-96 Writing Portfolio Audit Final Report

The 1996-97 Writing Portfolio Audit: Rationale and Procedures

The 1997-98 Writing Portfolio Audit: Rationale and Procedures

THE KIRIS MATHEMATICS PORTFOLIO

Since the Mathematics portfolio was discontinued in 1996 and was not included in the accountability system for the accountability years of Cycle 3, please refer to *the KIRIS Accountability Cycle 2 Technical Manual* for discussions of the rationale, design, professional development, monitoring and consistency of the last two years of the existence of the mathematics portfolio. Please refer to Chapter 7 for information about the two-year research and development phase.

THE KIRIS ALTERNATE PORTFOLIO

The Alternate Portfolio is the assessment vehicle for students with what in general terms can be described as moderate to severe disabilities that prevent them from participating in regular classroom instruction even after all possible assistance and instructional adaptations have been provided. Such circumstances likewise prevent student participation in the regular KIRIS testing program.

RATIONALE AND DESIGN

A team composed of teachers from the entire state, local administrators, Kentucky Department of Education staff, and University of Kentucky staff developed the Alternate Portfolio. The purpose was to reflect educational outcomes that were important for all students, including students with moderate to severe disabilities, and consistent with the Kentucky Education Reform Act. Eligibility for the alternate portfolio was based upon the student's need for an individualized curriculum based upon life-skill needs.

The Alternate Portfolio included a table of contents, a letter to the reviewer from the student, which could be written with the help of his/her teacher or non-disabled peers, a letter from the parent with their thoughts on the student's portfolio, and seven to ten entries. Portfolio entries included the student's written school schedule enhanced with photographs or pictures so that the student could understand it, examples of how the student communicated with others, a school and community job resume for the student if he or she were listed as a grade 8 or grade 12 student, and other entries selected and developed by the student.

The Alternate Portfolio was scored at grades 4, 8, and 12. For a student enrolled in a non-graded program, the portfolio was scored the year that the student was nine years

old as of October 1, 13 years old as of October 1 or during his/her last year in school.

The Alternate Portfolio employed a three-level moderation scoring system. Each student's teacher scored the portfolio, which was called the district scoring. Then, another teacher from within the same region of the state scored the portfolio. If they did not agree and the original teacher assigned score, was Proficient or Distinguished, the portfolio was re-scored at the state level to establish a final score. Any time there was a difference between district and regional scoring, a district was permitted to request that the portfolio be re-scored at the state level. Any Alternate Portfolio differing by two or more performance levels in the district and regional scoring was re-scored at the state level.

MONITORING THE SYSTEM

Scoring consistency was based on agreement between the original, district assigned score and the regional score assigned by a teacher who was generally from a different district. The scorers of Alternate Portfolios received training intended to be similar to that used for the writing portfolio. However, the number of scorers of Alternate Portfolios was much smaller than that needed for Writing Portfolios. Writing approximately 140,000 students developed Portfolios annually, compared to fewer than 1,000 Alternate Portfolios annually.

During Accountability Cycle 3, scoring accuracy was viewed as agreement between district and regional scores in that the regional scorers had undergone the same training but were not personally involved with the students developing the Alternate Portfolio. The agreement between either district or regional scores with scores assigned at the state level were not useful as an accuracy index because not all portfolios were re-scored at the state level and their selection was not a random process. Table 12-6 summarizes scoring consistency and accuracy based on a comparison of district and regional scores.

Both scoring consistency and accuracy declined for the Alternate Portfolio from 1993 to 1996. Exact agreement between local scores and regional scores dropped by 11 percent. During the same period, local scoring appears to have become less accurate for all grades assessed. Of the local scores changed by regional scorers in 1993, slightly more than half were lowered. By 1996, the number of scores lowered by regional scorers was roughly four times the number raised. Across all grades, the reduction amounted to 18 points on the 140-point KIRIS scale. The cause of the pattern is not clear. Scoring training and materials did not change substantially during this period.

More important than the differences in the delivery of training may be the difference in the opportunities a trained teacher is likely to have had to apply the training to real portfolios. Most teachers trained to score the Alternate Portfolio will encounter only one to five portfolios in a year and probably will not score Alternate Portfolios for accountability purposes every year. Observers report that teachers of students

qualifying for the Alternate Portfolio tend not to focus on the portfolio development (or at least the scoring) until the accountability year. That may mean that the standards were not well understood by the teacher until the accountability year and/or that entries prepared for the portfolio during non-accountability years did not reflect the standards.

REPORT ON ALTERNATE PORTFOLIOS AVAILABLE FROM KDE

The following table presents data concerning scoring consistency, accuracy, and percentage of recommended changes for the Alternate Portfolio.

TABLE 12-6 SUMMARY OF AGREEMENT BETWEEN DISTRICT AND REGIONAL SCORING OF ALTERNATE PORTFOLIOS				
	GRADES			
	All	4	8	12
YEAR	PERCENT EXACT AGREEMENT			
1993	63.4%	63.0%	68.0%	56.6%
1994	50.3%	55.6%	49.3%	44.8%
1995	52.2%	54.6%	52.0%	48.3%
1996	52.3%	50.0%	53.3%	53.9%
	PERCENT SCORES RAISED BY REGION			
1993	16.7%	21.4%	14.2%	14.3%
1994	12.5%	12.8%	11.8%	13.3%
1995	10.6%	11.7%	9.4%	11.0%
1996	9.5%	10.6%	9.0%	8.7%
	PERCENT SCORES LOWERED BY REGION			
1993	19.9%	15.7%	17.8%	29.1%
1994	37.2%	31.6%	39.0%	41.9%
1995	37.3%	33.7%	38.6%	40.7%
1996	38.3%	39.4%	37.7%	37.4%
	RATIO RAISED/LOWERED SCORES			
1993	0.84	1.37	0.80	0.49
1994	0.34	0.40	0.30	0.32
1995	0.29	0.35	0.24	0.27
1996	0.25	0.27	0.24	0.23
	DISTRICT ASSIGNED INDEX			
1993	22.7	18.9	20.1	32.1
1994	59.0	52.1	57.3	70.3
1995	66.8	66.3	61.6	75.1
1996	65.8	68.0	67.6	67.6

(Table Continued)

TABLE 12-6 (Continued) SUMMARY OF AGREEMENT BETWEEN DISTRICT AND REGIONAL SCORING OF ALTERNATE PORTFOLIOS				
	REGIONALLY ASSIGNED INDEX			
1993	21.1	21.8	18.5	24.3
1994	44.7	41.7	42.6	51.6
1995	50.6	52.6	44.4	56.8
1996	47.5	48.7	48.6	48.6

PORTFOLIO ISSUES SUMMARY

Despite the differences in scoring activities and portfolio types, the data present an interesting pattern in scoring accuracy across all three portfolios. Teachers in the accountability grades scored many Writing Portfolios each year and their agreement with external scorers was generally greater than 70 percent and improved over time. Moreover, they were faced every year with the challenge of shaping their instruction to align with portfolio scoring criteria. The Alternate Portfolios, which were infrequently encountered, were less robust in terms of scoring accuracy.

The connections between portfolio assessment, related professional development, and instructional practice have not yet been well documented. The fact that independent groups of trained readers draw the same conclusions about instructional program strengths and weakness after scoring a school set of portfolios is important. It shows one way for a school to use portfolio analysis information to plan relevant professional development and instruction for the coming year. When a standard procedure for portfolio analysis is used across schools, the state can begin to link instructional practices with improved portfolio scores.

Until teachers acknowledge that instructional change is the key to improving student performance, one of the major benefits of portfolio assessment will be unrealized and results unlikely to change. On the other hand, high gains in Writing Portfolio scores from several schools were validated by the 1996 audit. Many of these schools were smaller than average. Speculation about this fact was answered by teachers' explanation that small schools find it easier to adopt an intense school-wide focus on a particular program and, it is this planned alignment between assessment criteria, professional development, and instructional change that accounts for large portfolio score gains. KDE efforts to document the portfolio practices of high performing schools and to build that information into state wide portfolio training activities must continue so that all schools can effectively use portfolio assessment to document and improve student performance.

This page was intentionally left blank.